PRECISE ERROR ANALYSIS OF THE ℓ_2-LASSO

Christos Thrampoulidis[*] Ashkan Panahi[†], Daniel Guo[*], Babak Hassibi[*]

* Department of Electrical Engeeniring, Caltech, Pasadena, USA
† Signal Processing Group, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

A classical problem that arises in numerous signal processing applications asks for the reconstruction of an unknown, k-sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ from underdetermined, noisy, linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$. One standard approach is to solve the following convex program $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda\|\mathbf{x}\|_1$, which is known as the ℓ_2-LASSO. We assume that the entries of the sensing matrix $\mathbf{A}$ and of the noise vector $\mathbf{z}$ are i.i.d Gaussian with variances $1/m$ and σ^2. In the large system limit when the problem dimensions grow to infinity, but in constant rates, we *precisely* characterize the limiting behavior of the normalized squared error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\sigma^2$. Our numerical illustrations validate our theoretical predictions.

Index Terms— LASSO, square-root LASSO, normalized squared error, sparse recovery, Gaussian min-max theorem

1. INTRODUCTION

1.1. Motivation

The Least Absolute Shrinkage and Selection Operator (LASSO) is a celebrated convex progam used to estimate sparse signals from noisy linear underdetermined observations. Given a vector of observations $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$ of an unknown, but k-*sparse* (i.e., at most k nonzero entries), signal $\mathbf{x}_0 \in \mathbb{R}^n$, the ℓ_2-LASSO[1] produces the following estimate for $\mathbf{x}_0$:

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \Phi(\mathbf{x}; \mathbf{A}, \mathbf{z}) := \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \frac{\lambda}{\sqrt{m}}\|\mathbf{x}\|_1. \quad (1)$$

Here, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the sensing matrix, $\mathbf{z} \in \mathbb{R}^m$ is the noise vector and $\lambda > 0$ is a regularizer parameter. The LASSO has been long investigated from different perspectives and shown to enjoy unique properties, in terms of both computation and precision. Yet, some important asymptotic properties of it have not yet been fully understood. Our interest is on the *exact* characterization of the reconstruction error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2$.

1.2. Contribution

We assume a generic setup in which the entries of the sensing matrix and the *non-zero* entries of $\mathbf{x}_0$ are i.i.d Gaussian.

[1]Also known as "square-root LASSO" [1]. Please refer to [2, Sec. 1.3].

Also, the noise vector $\mathbf{z}$ is assumed to have entries i.i.d Gaussian with variance σ^2. Under this assumption we derive an *asymptotically exact* expression for the normalized squared error (NSE) $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2/\sigma^2$ of the ℓ_2L-ASSO. In the low noise regime $\sigma^2 \to 0$, our result reduces to simple and interpretable formulae. Although our theoretical analysis requires an asymptotic setting in which the problem dimensions grow to infinity, our numerical illustrations suggest that the predictions be accurate already with problem dimensions ranging over only a few hundreds. Also, we remark on our assumption on the Gaussian nature of the sensing matrix. This assumption has a long tradition in the statistics literature and sheds important insights [3]; the Gaussian ensemble has wonderful properties which make the analysis tractable, while at the same time many of the results generalize to a wider class of distributions. The main technical tool used in our analysis is a gaussian comparison inequality due to Gordon [4]. When combined with appropriate convexity assumptions, the inequality can be shown to be tight [2, 5, 6, 7], which makes it ideal for our precise analysis.

1.3. Relevant Literature

The LASSO was introduced by Tibshirani in [8] in the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \text{ s.t. } \|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1. \quad (2)$$

(1) is a regularized version of (2). An alternative to (1) solves

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_1. \quad (3)$$

Both this and (1) are variations of the same algorithm and Lagrange duality ensures that they both become equivalent to the constrained optimization (2) for proper choice of the regularizer parameters. There are reasons to argue in favor of either of them [1, 9], but they go beyond the scope of this paper. Early well-known bounds on the reconstruction error of the LASSO were order-wise in nature (i.e. accurate only up to constant multiplicative factors) and derived based on RIP and Restricted Eigenvalue assumptions on the measurement matrix [1, 10, 11, 12, 13]. To the best of our knowledge, the first precise asymptotic evaluation of the limiting behavior of the LASSO reconstruction error is due to Bayati and Montanari [3]; they consider problem (3), i.i.d Gaussian sensing matrix $\mathbf{A}$ and use the Approximate Message Passing (AMP) framework. More recently, and closer to our work, Stojnic

introduced an alternative framework that cleverly uses a celebrated gaussian comparison inequality due to Gordon [4]; he applied this on (2) and obtained precise characterizations for its worst-case NSE, which he showed to occur in the limit $\sigma^2 \to 0$. The same framework was used in [2] to generalize the results of [14] to arbitrary convex regularizer functions. Moreover, the authors in [2] consider (1) and obtain simple, precise error formulae when $\sigma^2 \to 0$. Our work, extends this result of [2] in several directions for the case of sparse recovery. First, it holds for arbitrary values of the noise-variance σ^2; to our knowledge, this is the first such precise asymptotic result for (1) ([3] considers (3)), and suggests the capabilities of the framework used. Also, when $\sigma^2 \to 0$, it recovers the result of [2] and extends it on the range of values of λ for which it holds. Finally, we believe that our result can be used to prove that the worst-case NSE of (1) occurs when $\sigma^2 \to 0$. This would imply that the simple and interpretable expressions corresponding to that regime are tight bounds on the NSE for arbitrary noise levels. Some additional effort is required to prove this claim and is, thus, left for future work.

2. ANALYSIS

2.1. Problem Setup

For the rest of the paper, let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution of mean μ and variance σ^2. Also, we use $\|\cdot\|$ instead of $\|\cdot\|_2$. Let $\mathbf{x}_0 \in \mathbb{R}^n$ denote the unknown signal and $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$ denote the vector of observations. We make the following assumptions:

- the entries of $\mathbf{A}$ are i.i.d $\mathcal{N}(0, 1/m)$,
- the entries of $\mathbf{z}$ are i.i.d $\mathcal{N}(0, \sigma^2)$,
- $\mathbf{x}_0$ is k-sparse with *fixed* support set S, i.e. $(\mathbf{x}_0)_i = 0$ for all $i \notin S$. Also, $(\mathbf{x}_0)_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$,

Consider estimating $\mathbf{x}_0$ via the solution $\hat{\mathbf{x}} := \hat{\mathbf{x}}(\lambda)$ of the LASSO program in (1). Our goal is to provide tight expressions for the reconstruction error $\|\hat{\mathbf{x}} - \mathbf{x}_0\|$; this depends explicitly on $\mathbf{x}_0$ and implicitly on $\mathbf{A}$ and $\mathbf{z}$. We consider an asymptotic setting in which the problem parameters n, m and k grow proportionally as $m/n \to \delta \in (0, 1)$ [2] and $k/m \to \gamma \in (0, 1)$. Our result characterizes the limiting behavior of the quantity of interest as $n \to \infty$. To suppress notation, we use the symbol "$\approx$" to denote convergence in probability, i.e. $X_n \approx X$ is used to denote that a sequence of random variables X_n converges in X in that for all $\epsilon > 0$, $\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \epsilon X) = 0$. Similarly, we write $X_n \gtrsim X$ if for all $\epsilon > 0$, $\lim_{n \to \infty} \mathbb{P}(X_n < (1 - \epsilon)X) = 0$.

2.2. Preliminaries

It is convenient to rewrite (1) changing the optimization variable to the error vector $\mathbf{w} := \mathbf{x} - \mathbf{x}_0$: [3]

$$\min_{\mathbf{w}} \Phi(\mathbf{w}; \mathbf{A}, \mathbf{z}) := \sqrt{m}\|\mathbf{z} - \mathbf{A}\mathbf{w}\| + \lambda\|\mathbf{x}_0 + \mathbf{w}\|_1. \quad (4)$$

Denote $\hat{\mathbf{w}}^{\lambda,\sigma}(\mathbf{A}, \mathbf{z})$ the solution of (4). We often drop the dependence on the arguments $\lambda, \sigma, \mathbf{A}, \mathbf{z}$ when clear from context. Suppose we want to show that $\|\hat{\mathbf{w}}\| \approx \alpha_*$, for appropriate $\alpha_* \geq 0$. Equivalently, for all $\epsilon > 0$ and sets $\mathcal{R}_\epsilon := \{\ell \mid |\ell - \alpha_*| > \epsilon\alpha_*\}$, we wish that $\lim_{n \to \infty} \mathbb{P}(\|\hat{\mathbf{w}}\| \in \mathcal{R}_\epsilon) = 0$. It suffices to prove the existence of $\delta := \delta(\epsilon) > 0$ such that

$$\lim_{n \to \infty} \mathbb{P}\left(\min_{\|\mathbf{w}\| \in \mathcal{R}_\epsilon} \Phi(\mathbf{w}) \leq (1 + \delta)\Phi(\hat{\mathbf{w}})\right) = 0. \quad (5)$$

Directly showing (5) through analyzing $\Phi(\mathbf{w})$ is difficult. Instead, we use the Gaussian min-max Theorem to translate the LASSO objective function in (4) to a simpler one, that is amenable to direct analysis.

2.3. Introducing a Simpler Optimization

The Gaussian min-max Theorem belongs to the family of the so called Gaussian comparison inequalities [15, Ch. 3] and was proved by Gordon in [4, Lemma 3.1]. To see how that theorem can be applied in our case, observe that

$$\Phi(\mathbf{w}) = \max_{\|\mathbf{u}\| \leq 1} \mathbf{u}^T[\sqrt{m}\mathbf{A}, -\mathbf{z}/\sigma]\begin{bmatrix}\mathbf{w} \\ \sqrt{m}\sigma\end{bmatrix} + \lambda\|\mathbf{x}_0 + \mathbf{w}\|_1.$$

Let $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{h} \in \mathbb{R}^n$, $g \in \mathbb{R}$ have i.i.d $\mathcal{N}(0, 1)$ entries and

$$\phi(\mathbf{w}; \mathbf{g}, \mathbf{h}) := \max_{\|\mathbf{u}\| \leq 1}\left\{\sqrt{\|\mathbf{w}\|^2 + m\sigma^2}\mathbf{g}^T\mathbf{u}\right.$$
$$\left. - \|\mathbf{y}\|(\mathbf{h}^T\mathbf{w} + \sqrt{m}\sigma g) + \lambda\|\mathbf{w} + \mathbf{x}_0\|_1\right\}. \quad (6)$$

Then, the Gaussian min-max Theorem[4] states that for arbitrary set $\mathcal{R} \subset \mathbb{R}^n$ and $c \in \mathbb{R}$:

$$\mathbb{P}(\min_{\mathbf{w} \in \mathcal{R}} \Phi(\mathbf{w}; \mathbf{A}, \mathbf{z}) \leq c) \leq 2\mathbb{P}(\min_{\mathbf{w} \in \mathcal{R}} \phi(\mathbf{w}; \mathbf{g}, \mathbf{h}) \leq c). \quad (7)$$

Recently, it was shown in [7, 14] that when combined with appropriate convexity assumptions, the Gaussian min-max Theorem is *tight*. In particular, for a *convex* compact set $\mathcal{R}$ it is also true [7, Theorem II.1] that

$$\mathbb{P}(\min_{\mathbf{w} \in \mathcal{R}} \Phi(\mathbf{w}; \mathbf{A}, \mathbf{z}) \geq c) \leq 2\mathbb{P}(\min_{\mathbf{w} \in \mathcal{R}} \phi(\mathbf{w}; \mathbf{g}, \mathbf{h}) \geq c). \quad (8)$$

(7) and (8) are critical for establishing (5). They suggest the analysis of the following optimization problem[5], which we refer to as "Gordon's optimization (GO)":

$$(GO) \qquad \phi(\mathbf{w}_*; \mathbf{g}, \mathbf{h}) := \min_{\mathbf{w}} \phi(\mathbf{w}; \mathbf{g}, \mathbf{h}), \quad (9)$$

in place of the original LASSO optimization in (4). To see how this can be useful, assume that $\phi(\mathbf{w}_*) \approx \mathcal{D}_*$; then, it would follow directly from (7) and (8) that $\Phi(\hat{\mathbf{w}}) \approx \mathcal{D}_*$. Next, we analyze such asymptotic properties of (GO) .

---

[2] Our analysis extends to the overdetermined case, where $\delta \in [1, \infty)$. For simplicity, in this paper we focus on the underdetermined regime.

[3] Also, note the re-scaling in (4) with a factor of $\sqrt{m}$.

[4] In fact, the statement in (7) is a slight variation of the original statement of the Gaussian min-max Theorem. Please refer to [7] for details.

[5] The proof of (8) in [7] requires the set $\mathcal{R}$ to be compact. This technical detail can be resolved by assuming a sufficiently large upper bound K on $\|\hat{\mathbf{w}}\|$ such that constraining the minimization in (9) over the set $\|\mathbf{w}\| \leq K$ does not change the optimal cost.

$$\mathbf{D}(\lambda) := k(1+\lambda^2) + (n-k)\rho(1,\lambda), \quad \mathbf{C}(\lambda) := -(\lambda/2)\partial\mathbf{D}(\lambda)/\partial\lambda, \quad \psi(\lambda) := \lambda/Q^{-1}\left(1/2 + (m - \mathbf{D}(\lambda) - \mathbf{C}(\lambda))/(2k)\right),$$

$$f(\lambda,\phi) := m - \mathbf{D}(\lambda) + m\sigma^2(1-\phi^2) + k\left(\phi^2 - \lambda^2 - 2\right)\left(2Q\left(\lambda/\phi\right)-1\right) + k\sqrt{2/\pi}\lambda\phi\exp\left(-\lambda^2/(2\phi^2)\right). \qquad (10)$$

2.4. Analyzing Gordon's Optimization

2.4.1. Scalarization

We begin with simplifying (GO) and reducing it to an optimization problem involving only scalars. For now, assume that $\mathbf{g}, \mathbf{h}$ and g are all fixed. Note that the maximization over the direction of $\mathbf{u}$ in (6) is easy to evaluate. Furthermore, we may write $\|\mathbf{x}_0+\mathbf{w}\|_1 = \max_{\|\mathbf{v}\|_\infty \leq 1}(\mathbf{x}_0+\mathbf{w})^T\mathbf{v}$. With these:

$$\phi(\mathbf{w}_*) = \min_{\mathbf{w}} \max_{\substack{0\leq\beta\leq 1 \\ \|\mathbf{v}\|_\infty\leq 1}} \left\{ \sqrt{\|\mathbf{w}\|^2 + m\sigma^2}\|\mathbf{g}\|\beta \right. \qquad (11)$$
$$\left. - \beta(\mathbf{h}-\lambda\mathbf{v})^T\mathbf{w} - \beta\sqrt{m}\sigma g + \lambda\mathbf{x}_0^T\mathbf{v}\right\}.$$

The objective function in (11) is convex-concave in $\mathbf{w}$ and $\beta, \mathbf{v}$. Also, the constraint sets are closed convex and one of them is bounded; thus, we can switch the corresponding order of min-max [16, Cor. 37.3.2]. After this, the minimization over the direction of $\mathbf{w}$ is easy to evaluate:

$$\phi(\mathbf{w}_*) = \max_{\substack{0\leq\beta\leq 1 \\ \|\mathbf{v}\|_\infty\leq 1}} \min_{\alpha\geq 0} \left\{ \sqrt{\alpha^2 + m\sigma^2}\|\mathbf{g}\|_2\beta - \beta\sqrt{m}g\sigma \right.$$
$$\left. - \alpha\|\lambda\mathbf{v} - \beta\mathbf{h}\|_2 + \lambda\mathbf{x}_0^T\mathbf{v}\right\}.$$

Note that the optimization variable α above plays the role of the ℓ_2-norm of $\mathbf{w}$; its optimal value α_* is equal to $\|\mathbf{w}_*\|_2$. Switching one last time the order of optimization between $\mathbf{v}, \beta$ and α, we conclude with the following optimization:

$$\min_{\alpha\geq 0} \max_{0\leq\beta\leq 1} \left\{ \sqrt{\alpha^2 + m\sigma^2}\|\mathbf{g}\|_2\beta - \beta\sqrt{m}g\sigma - \beta\mathcal{F}(\alpha,\beta)\right\},$$

where $\mathcal{F}(\alpha,\beta) := \min_{\|\mathbf{v}\|_\infty\leq 1}\left\{\alpha\|\frac{\lambda}{\beta}\mathbf{v} - \mathbf{h}\|_2 - \frac{\lambda}{\beta}\mathbf{x}_0^T\mathbf{v}\right\}$. The final step amounts to reducing the optimization over $\mathbf{v}$ to a scalar optimization. The idea here is to apply the fact that for any real $r \geq 0$, one has $\sqrt{r} = \min_{p>0}\left(\frac{p}{2} + \frac{r}{2p}\right)$ in order to make the objective function separable. With this:

$$\mathcal{F}(\alpha,\beta) = \min_{\|\mathbf{v}\|_\infty\leq 1} \min_{p>0} \frac{\alpha p}{2} + \frac{a\|\lambda/\beta\mathbf{v} - \mathbf{h}\|_2^2}{2p} - \frac{\lambda}{\beta}\mathbf{x}_0^T\mathbf{v}$$

$$= \min_{p>0} \left\{ \frac{\alpha p}{2} + \sum_{i\in S}\left[\frac{\alpha}{2p}\min_{|\mathbf{v}_i|\leq\frac{\lambda}{\beta}}\left(\mathbf{v}_i - \mathbf{h}_i - \frac{p(\mathbf{x}_0)_i}{\alpha}\right)^2\right.\right.$$

$$\left.\left. - \frac{p(\mathbf{x}_0)_i^2}{2\alpha} - \mathbf{h}_i(\mathbf{x}_0)_i\right] + \frac{\alpha}{2p}\sum_{i\notin S}\min_{|\mathbf{v}_i|\leq\frac{\lambda}{\beta}}(\mathbf{v}_i - \mathbf{h}_i)^2\right\}. \quad (12)$$

The second equality above follows from standard completion of squares. The scalar minimizations over $\mathbf{v}_i$'s in (12) are simple soft-thresholding operations: $\min_{|q|\leq\tau}(r-q)^2 = ((|r| - \tau)^+)^2$. Combining all the above, we have shown that $\phi(\mathbf{w}_*) = \min_{\alpha\geq 0}\max_{0\leq\beta\leq 1, p>0}\phi_o(\alpha,\beta,p)$, for an appropriately defined objective function $\phi_o(\alpha,\beta,p)$.

2.4.2. Concentration

Our next step is to analyze the limiting behavior of $\phi_o(\alpha,\beta,p)$. Recall that $\mathbf{g}, \mathbf{h}, g$ have i.i.d $\mathcal{N}(0,1)$ entries. Thus, $\|\mathbf{g}\| \approx \sqrt{m}$ and $g \approx 0$. Using these and applying the Law of Large Numbers to the summations in (12) it can be shown that $\phi_o(\alpha,\beta,p) \approx \mathcal{D}(\alpha,\beta,p)$, where

$$\mathcal{D}(\alpha,\beta,p) := \beta\left(\sqrt{\alpha^2 + m\sigma^2}\sqrt{m} - \frac{\alpha p}{2} + k\frac{p}{2\alpha}\right.$$
$$\left. - k\frac{\alpha}{2p}\rho(\sqrt{1 + p^2/\alpha^2}, \lambda/\beta) - (n-k)\frac{\alpha}{2p}\rho(1, \lambda/\beta)\right),$$

$$\rho(c,\tau) := \mathbb{E}_{g\sim\mathcal{N}(0,1)}\left[(|cg| - \tau)^+\right]^2$$
$$= 2(c^2 + \tau^2)Q(\tau/c) - \sqrt{2/\pi}c\tau e^{-\tau^2/(2c^2)}, \quad (13)$$

and $Q(x) = (1/\sqrt{2\pi})\int_x^\infty e^{-x^2/2}\mathrm{d}x$. Albeit some technicalities involved (skipped due to space limitations):

$$\min_{\mathbf{w}} \phi(\mathbf{w}) \approx \min_{\alpha\geq 0}\max_{\substack{0\leq\beta\leq 1 \\ p>0}} \mathcal{D}(\alpha,\beta,p), \qquad (14a)$$

$$\min_{\|\mathbf{w}\|\in\mathcal{R}_\epsilon} \phi(\mathbf{w}) \approx \min_{\alpha\in\mathcal{R}_\epsilon}\max_{\substack{0\leq\beta\leq 1 \\ p>0}} \mathcal{D}(\alpha,\beta,p). \qquad (14b)$$

with $\mathcal{R}_\epsilon = \{\ell \mid |\ell - \alpha_*| > \epsilon\alpha_*\}$ and α_* minimizer in (14a).

2.5. Back from Gordon's Optimization to the LASSO

Once we have analyzed (GO), we may now appropriately apply the Gaussian min-max Theorem (in particular, [7, Thm. II.1]) to conclude with the following result.

Theorem 2.1 (LASSO Objective). *Recall the definitions of $\Phi(\mathbf{w})$, $\mathcal{D}(\alpha,\beta,p)$ in (4) and (13), respectively. Also, let $\Phi_* := \min_{\mathbf{w}}\Phi(\mathbf{w})$ and $\mathcal{D}_* := \min_{\alpha\geq 0}\max_{\substack{0\leq\beta\leq 1 \\ p>0}}\mathcal{D}(\alpha,\beta,p)$. Then, for all $\epsilon > 0$, $\lim_{n\to\infty}\mathbb{P}(|\Phi_* - \mathcal{D}_*| > \epsilon\mathcal{D}_*) = 0$.*

The proof follows from (14a) when combining (7) and (8). Also, a single application of (7) in (14b) shows that for any $\mathbb{R}_\epsilon \subset \mathbb{R}$:

$$\Phi_\epsilon := \min_{\|\mathbf{w}\|\in\mathcal{R}_\epsilon} \Phi(\mathbf{w}) \gtrsim \min_{\alpha\in\mathcal{R}_\epsilon}\max_{\substack{0\leq\beta\leq 1 \\ p>0}} \mathcal{D}(\alpha,\beta,p) =: \mathcal{D}_\epsilon. \quad (15)$$

To appreciate the power of Theorem 2.1, note that it gives an explicit evaluation of the limiting behavior of the LASSO optimal cost in terms of the optimal cost of a much simpler, scalar and deterministic optimization problem. What is more, we can combine Theorem 2.1 with (15), to prove the following result about the limiting behavior of the LASSO error.

Theorem 2.2 (LASSO error). *Let $\hat{\mathbf{x}}$ be a minimizer of the LASSO in (1) and recall the definition of function $\mathcal{D}(\alpha,\beta,p)$ (13). If α_* is optimal for $\min_{\alpha\geq 0}\max_{\substack{0\leq\beta\leq 1 \\ p>0}}\mathcal{D}(\alpha,\beta,p)$, then for all $\epsilon > 0$: $\lim_{n\to\infty}\mathbb{P}(|\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2 - \alpha_*| > \epsilon\alpha_*) = 0$.*

We provide a sketch of the proof here (see also [7, Thm. II.1]). Consider the event

$$\mathcal{E} = \{\Phi_* \le (1+\zeta_1)\mathcal{D}_* \text{ and } \Phi_\epsilon \ge (1-\zeta_2)\mathcal{D}_\epsilon\}.$$

In view of Thm. 2.1 and (15): $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}) = 1$. Hence,

$$\lim_{n\to\infty} \mathbb{P}(\Phi_\epsilon \le (1+\delta)\Phi_*) \le \lim_{n\to\infty} \mathbb{P}(\Phi_\epsilon \le (1+\delta)\Phi_*|\mathcal{E})$$
$$\le \mathbb{P}((1-\zeta_2)\mathcal{D}_\epsilon \le (1+\delta)(1+\zeta_1)\mathcal{D}_*). \quad (16)$$

Next, we show that we can choose ζ_1, ζ_2, δ such that the (deterministic) event in (16) does not occur; this will complete the proof of (5). It can be shown that the function $\mathcal{D}(\cdot, \beta, p)$ is strictly convex for all β, p. Thus, $\exists$ constant $L > 0$:

$$\mathcal{D}_\epsilon - \mathcal{D}_* =: \mathcal{D}(\alpha_\epsilon, \beta_\epsilon, p_\epsilon) - \mathcal{D}(\alpha_*, \beta_*, p_*) =$$
$$\mathcal{D}(\alpha_\epsilon, \beta_\epsilon, p_\epsilon) - \mathcal{D}(\alpha_\epsilon, \beta_*, p_*) + \mathcal{D}(\alpha_\epsilon, \beta_*, p_*) - \mathcal{D}(\alpha_*, \beta_*, p_*)$$
$$\ge 0 + L|\alpha_\epsilon - \alpha_*|^2 \ge \epsilon^2 L \alpha_*^2. \quad (17)$$

Denote $\theta = \theta(\epsilon) := \epsilon^2 L \alpha_*^2 / \mathcal{D}_*$. Set $\zeta_1 = \theta/4$, $\zeta_2 = \frac{\theta/2}{1+\theta}$, $\delta = \frac{\theta}{4+\theta}$ and apply (17) to complete the proof.

2.6. Simplifying the Result

For any values of λ and σ, we ask for an accurate prediction of the LASSO error $\|\hat{\mathbf{x}}^{\lambda,\sigma} - \mathbf{x}_0\|$. Theorem 2.2 yields an answer as the solution of a minimax optimization, which is scalar, deterministic and convex. We have put some additional effort in order to simplify this optimization and make it more explicit. In particular, we can show that it has a unique global optimal solution, the evaluation of which essentially breaks down to solving two nonlinear one-dimensional algebraic equations (see Theorem 3.1). Apart from the computational advantage of this alternative description, it also allows for further theoretical insights. For instance, starting from Theorem 3.1 we believe that it is possible to prove that the worst-case NSE is attained in the limit of the noise variance $\sigma^2 \to 0$.

3. RESULTS

3.1. Arbitrary SNR Values

For the statement of our main result, recall (13) and further consider the definitions in (10). We also need the following.

Definition 3.1 (λ_{crit}^σ). *Define λ_{crit}^σ as the unique solution of the equation $f(x, \frac{x}{\psi(x)}) = 0$. Further, define $\lambda_{crit}^0 :=$*
$$Q^{-1}\left(\frac{1}{2}\frac{m-k}{n-k}\right) \text{ and }$$
$$\lambda_{crit}^{min} := \begin{cases} Q^{-1}\left(\frac{1}{2}\frac{m}{n-k}\right) & \text{if } m < n-k, \\ 0 & \text{else.} \end{cases}$$

It can be shown that for all $\sigma > 0$: $\lambda_{crit}^{min} \le \lambda_{crit}^\sigma \le \lambda_{crit}^0$.

Theorem 3.1. *Let $\hat{\mathbf{x}}^{\lambda,\sigma}$ be the solution of the LASSO in (1). Recall the definitions of f and λ_{crit}^σ in (10) and Definition 3.1. Let $\hat{\lambda} := \max\{\lambda, \lambda_{crit}^\sigma\}$ and $q_*(\alpha) = \sqrt{1 + \frac{m}{\sigma^2(m+\alpha^2)}}$. Denote α_* the unique solution of the equation $f(\hat{\lambda}, q_*(\alpha)) = 0$ with respect to α. Then, the following limit holds in probability: $\lim_{n\to\infty} \frac{\|\hat{\mathbf{x}}^{\lambda,\sigma} - \mathbf{x}_0\|_2}{\sigma\sqrt{m}} = \alpha_*$.*

3.2. Asymptotic NSE

Here we consider the case $\sigma^2 \to 0$.

Theorem 3.2. *Let $\hat{\lambda} := \max\{\lambda, \lambda_{crit}^0\}$. If $m \ge \mathcal{D}(\hat{\lambda})$, then the following limit holds in probability*
$$\lim_{n\to\infty} \lim_{\sigma\to 0} \frac{\|\hat{\mathbf{x}}^{\lambda,0} - \mathbf{x}_0\|_2}{\sigma\sqrt{m}} = \sqrt{\frac{\mathbf{D}(\hat{\lambda})}{m - \mathbf{D}(\hat{\lambda})}}.$$

When $\lambda > \lambda_{crit}$, Theorem 3.2 recovers the result of [2, Theorem 3.2]; adding to this, we are able to characterize the behavior when $\lambda < \lambda_{crit}$. The theorem suggests that stable recovery is possible only when $m > \mathcal{D}(\hat{\lambda})$. In particular, we require that at least $m > \min_{\lambda>0}\mathbf{D}(\lambda)$. A recent line of work [17, 18, 19, 20], has shown that $\min_{\lambda>0}\mathbf{D}(\lambda)$ precisely characterizes the minimum number m of required measurements for exact recovery of sparse signals under *noiseless* linear measurements. Also, the minimum of the formula that appears in the theorem is achieved at $\lambda_{best} := \arg\min_{\lambda>0}\mathbf{D}(\lambda)$. We refer the reader to Figure 1 for an illustration of $\lambda_{crit}^0, \lambda_{best}$ and to [2, Section 4.2] for a detailed further discussion.

4. SIMULATION RESULTS AND CONCLUSION
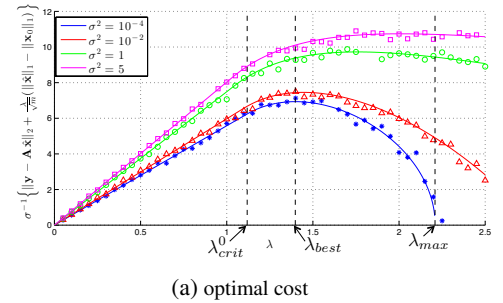
(a) optimal cost
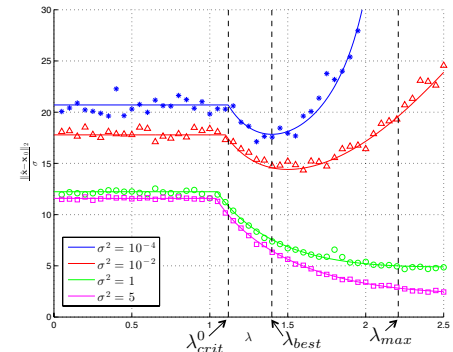
(b) normalized error

Fig. 1: $n = 500$, $m = 150$, $k = 20$. Averages over 20 realizations.

In Figure 1, observe the close agreement of the simulation results to the predictions of Theorems 3.1 and 3.2 (for the case $\sigma^2 = 10^{-4}$, we used Theorem 3.2 for the prediction). Refer to Section 3.2 for the definitions of $\lambda_{crit}^0, \lambda_{best}$; λ_{max} is such that $m < \mathbf{D}(\lambda)$ for all $\lambda > \lambda_{max}$. Our expressions are accurate for problem sizes on the order of a few hundreds and valid for all values of σ^2. In Figure 1b, the worst-case NSE occurs in the small noise-variance regime. We believe that this claim can be proved using Theorems 3.1 and 3.2.

References

[1] A. Belloni, V. Chernozhukov, and L. Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[2] S. Oymak, C. Thrampoulidis, and B. Hassibi, "The squared-error of generalized lasso: A precise analysis," *arXiv preprint arXiv:1311.0830*, 2013.

[3] M. Bayati and A. Montanari, "The lasso risk for gaussian matrices," *Information Theory, IEEE Transactions on*, vol. 58, no. 4, pp. 1997–2017, 2012.

[4] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$*. Springer, 1988.

[5] M. Stojnic, "Meshes that trap random subspaces," *arXiv preprint arXiv:1304.0003*, 2013.

[6] C. Thrampoulidis and B. Hassibi, "Estimating structured signals in sparse noise: A precise noise sensitivity analysis," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*. IEEE, 2014, pp. 866–873.

[7] C. Thrampoulidis, S. Oymak, and B. Hassibi, "A tight version of the gaussian min-max theorem in the presence of convexity," *arXiv preprint arXiv:1408.4837*, 2014.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[9] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Simple error bounds for regularized noisy linear inverse problems," *Information Theory, 2014. ISIT 2014. Proceedings. International Symposium on*.

[10] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[11] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, pp. 2313–2351, 2007.

[12] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *The Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.

[13] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers," *Statistical Science*, vol. 27, no. 4, pp. 538–557, 2012.

[14] M. Stojnic, "A framework to characterize performance of lasso algorithms," *arXiv preprint arXiv:1303.7291*, 2013.

[15] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes*. Springer, 1991, vol. 23.

[16] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1997, vol. 28.

[17] M. Stojnic, "Various thresholds for ℓ_1-optimization in compressed sensing," *arXiv preprint arXiv:0907.3666*, 2009.

[18] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[19] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: A geometric theory of phase transitions in convex optimization," *arXiv preprint arXiv:1303.6672*, 2013.

[20] M. Stojnic, "Upper-bounding ℓ_1-optimization weak thresholds," *arXiv preprint arXiv:1303.7289*, 2013.